Information Visualization

SWE 632, Spring 2018

GEORGE MASON UNIVERSITY

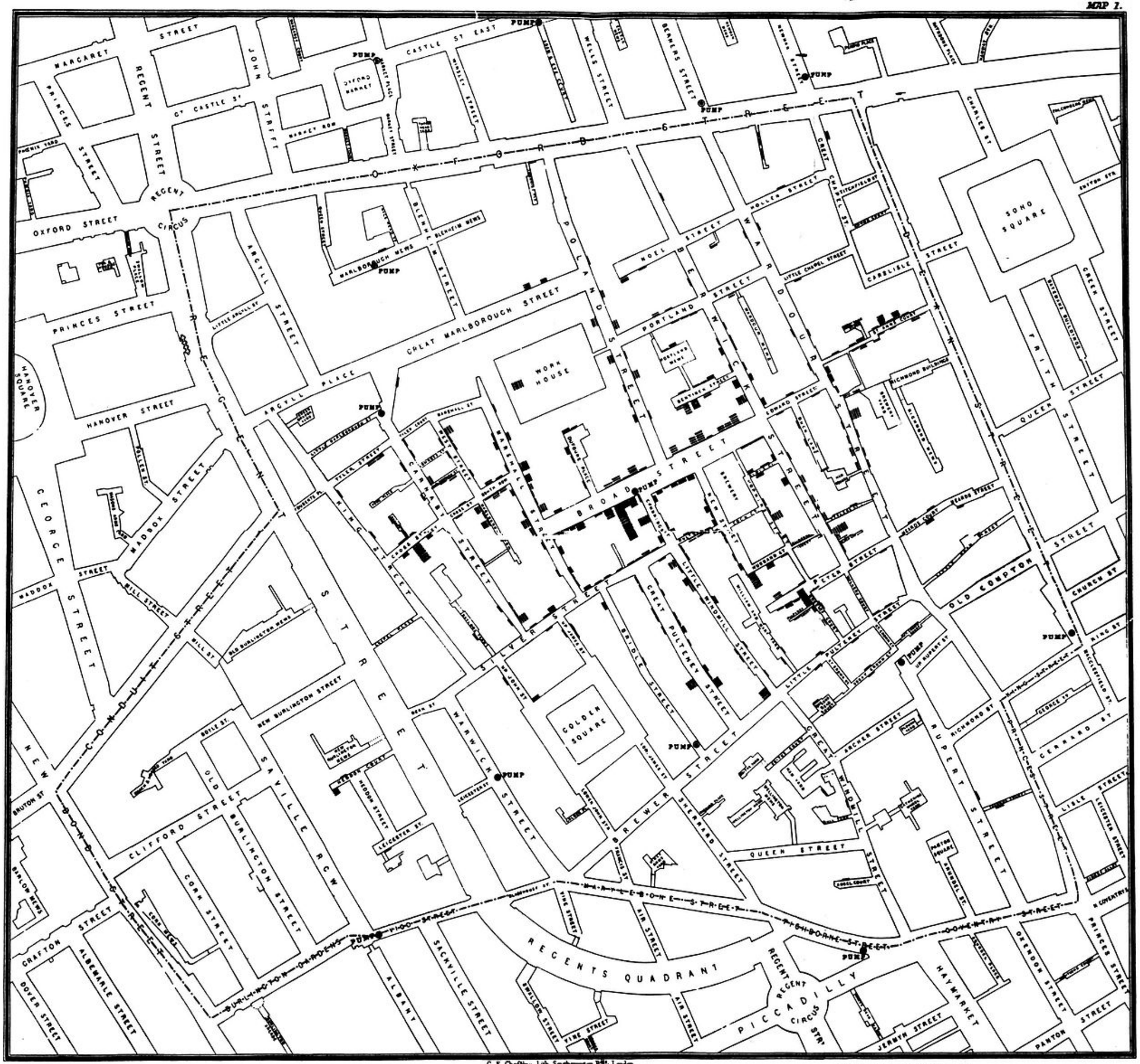
Today

- What types of information visualization are there?
  - Which one should you choose?
- What principles and guidelines inform the design of information visualizations?
- How can interactivity be used to design better information visualizations?

Cholera Epidemic in London, 1854

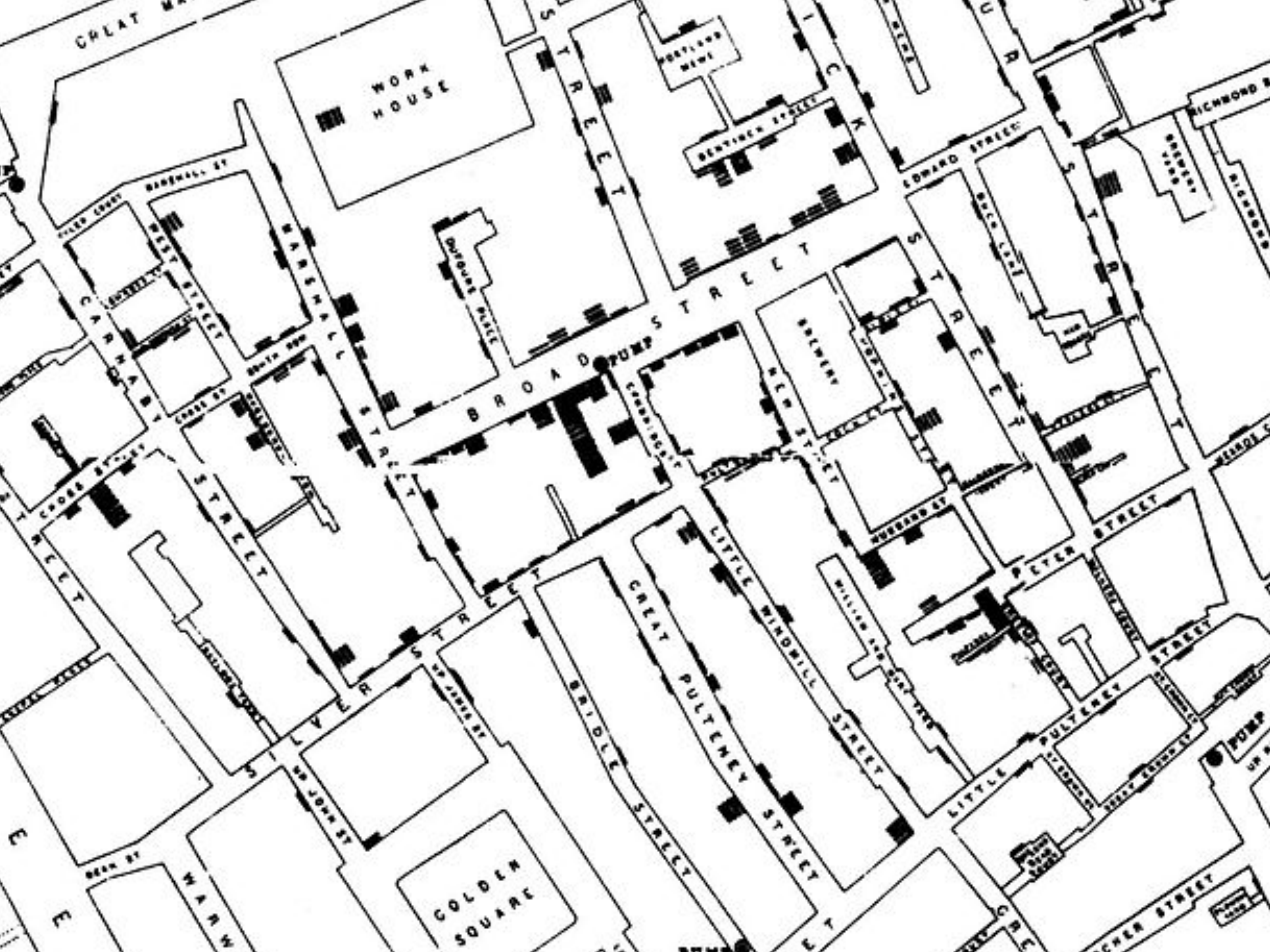
- >500 fatal attacks of cholera in 10 days
  - Concentrated in Broad Street area of London
  - Many died in a few hours
- Dominant theory of disease: caused by noxious odors
- Afflicted streets deserted by >75% inhabitants

MAP 1.



C. F. Cheffins, Lith. Southampton Bldgs. London.

SCALE 30 INCHES TO A MILE.


WORK HOUSE
BROAD STREET
PUMP
PORTLAND MEWS
BENTINCK STREET
EDWARD STREET
RICHMOND
BREWERY
NEW STREET
HUBBARD ST
PETER STREET
MARSHALL ST
CARNABY
WEST STREET
CROSS STREET
SILVER STREET
BRIDLE STREET
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
LITTLE PULTENEY
GOLDEN SQUARE
CAMBRIDGE ST
WARW

Investigation and aftermath

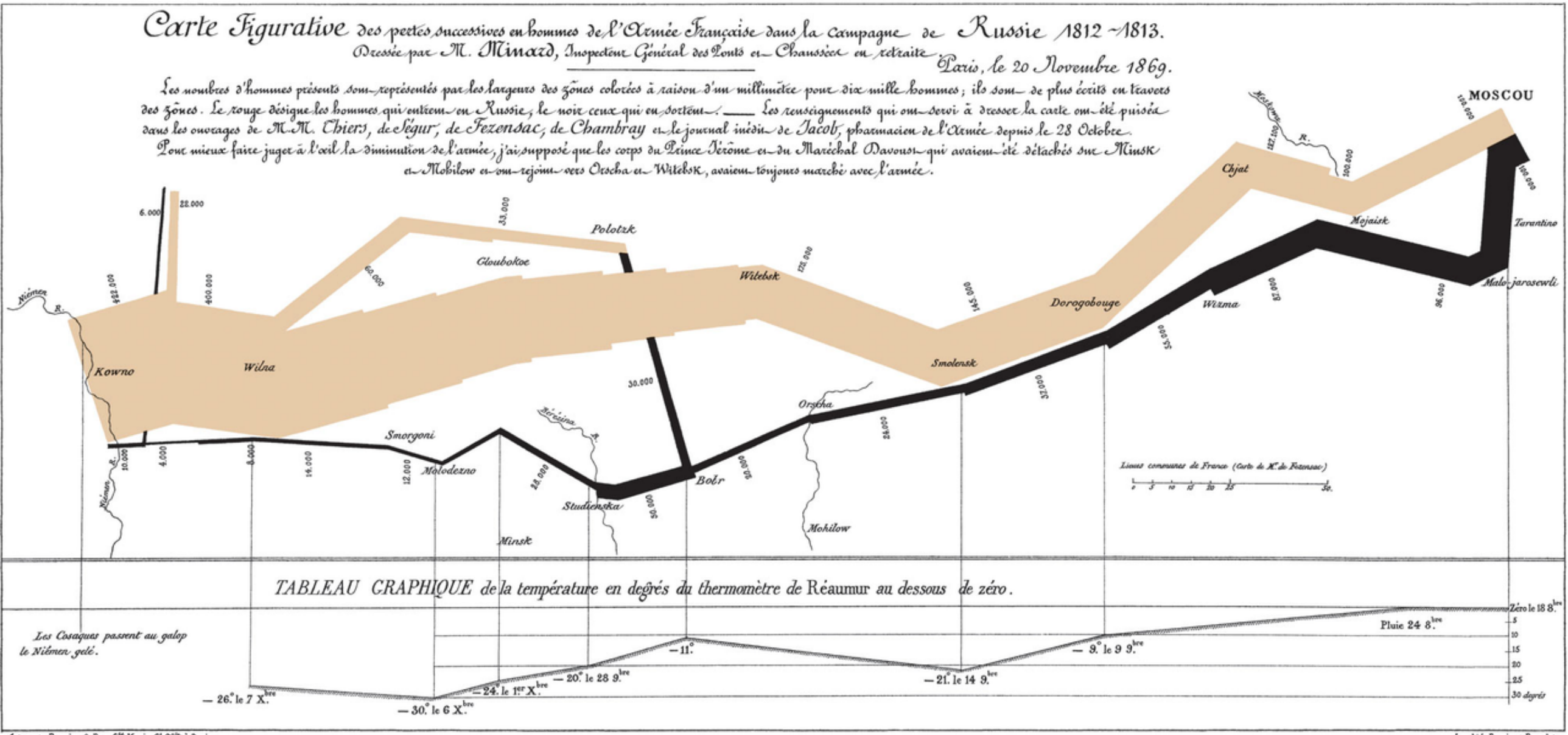
- Based on **visualization**, did case by case investigation
- Found that **61 / 83** positive identified as using well water from Broad Street pump
- Board ordered pump-handle to be removed from well
- Epidemic soon **ended**
- Solved centuries old question of how cholera spread

Methods used by Snow

- Placed data in appropriate **context** for assessing cause & effect
  - Plotted on map, included well location
  - Reveals proximity as cause
- Made quantitative **comparisons**
  - Fewer deaths closer to brewery, could investigate cause
- Considered **alternative** explanations & contrary cases
  - Investigated cases not close to pump, often found connection to pump
- Assessment of possible **errors** in numbers

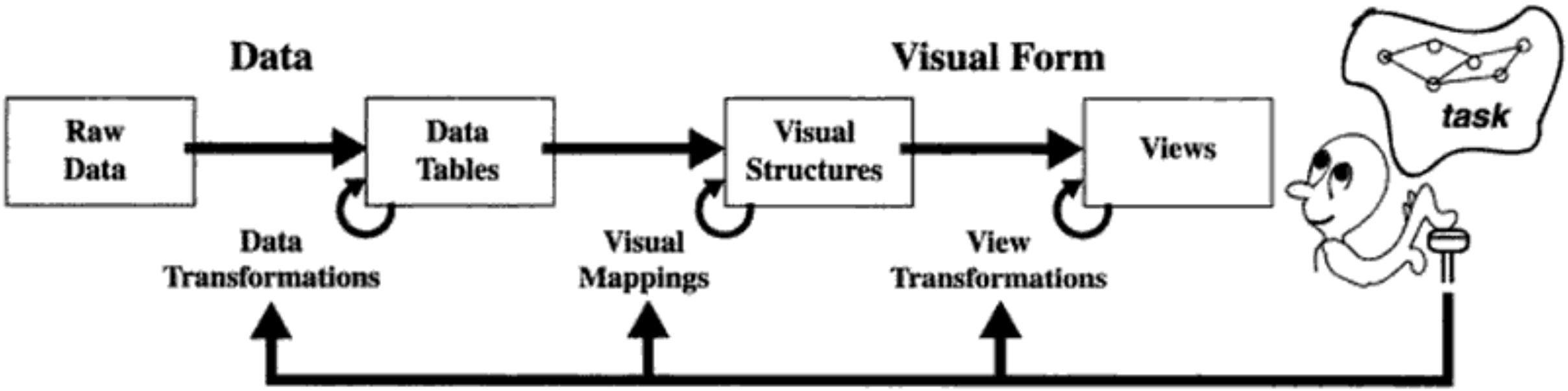
Charles Minard's Map of Napoleon's Russian Campaign of 1812

Drawn in Paris, November 20, 1869.



Mapping data to visual form

Designing an information visualization



Raw Data: idiosyncratic formats
Data Tables: relations (cases by variables) + metadata
Visual Structures: spatial substrates + marks + graphical properties
Views: graphical parameters (position, scaling, clipping. . . .)

Types of raw data

- Nominal - unordered set **without** a quantitative value
  - Gender: male, female
  - Hair color: brown, black, blonde, gray, orange, ...
- Ordinal - **ordered** set, with no meaning assigned to differences
  - How do you feel today: very unhappy, unhappy, ok, happy, very happy
  - Undefined how much better happy is than ok
- Quantitative - **numeric** value
  - Height, weight, distance, ...

Data transformations

- Classing / binning: Quantitative —> ordinal
  - Maps ranges onto **classes** of variables
  - Can also count # of items in each class w/ histogram
- Sorting: Nominal —> ordinal
  - Add order between items in sets
- Descriptive statistics: mean, average, median, max, min, …

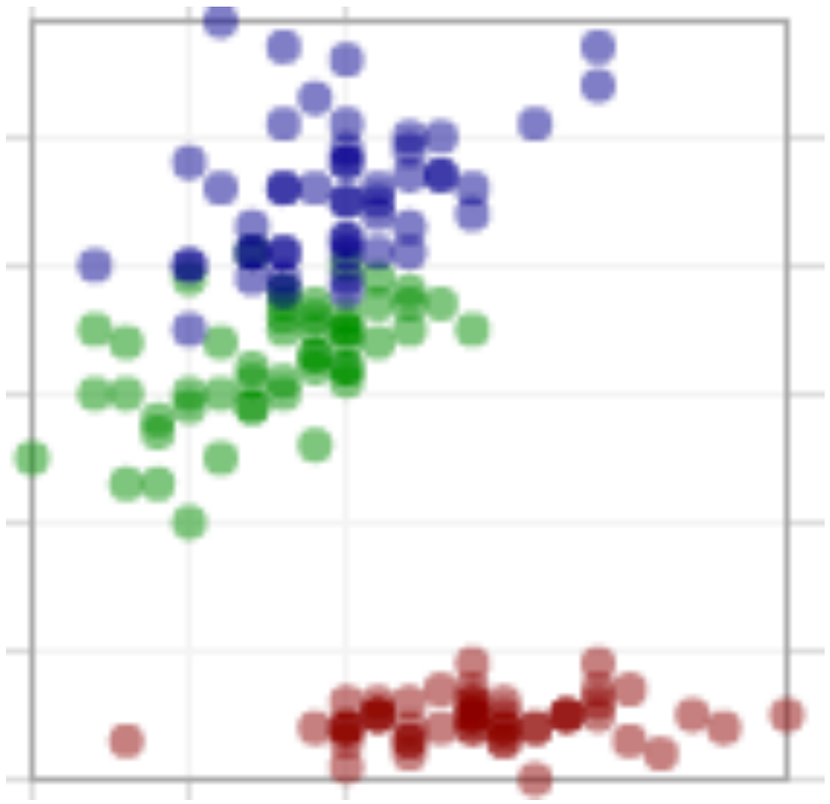
Example uses of a data transformation

Visual structures

- 3 components
  - spatial substrate
  - marks
  - marks' graphical properties

Spatial substrate

- Axes that divide space
- Types of axes - unstructured, nominal, ordinal, quantitative
- Composition - use of multiple orthogonal axes (e.g., 2D scatterplot, 3D)

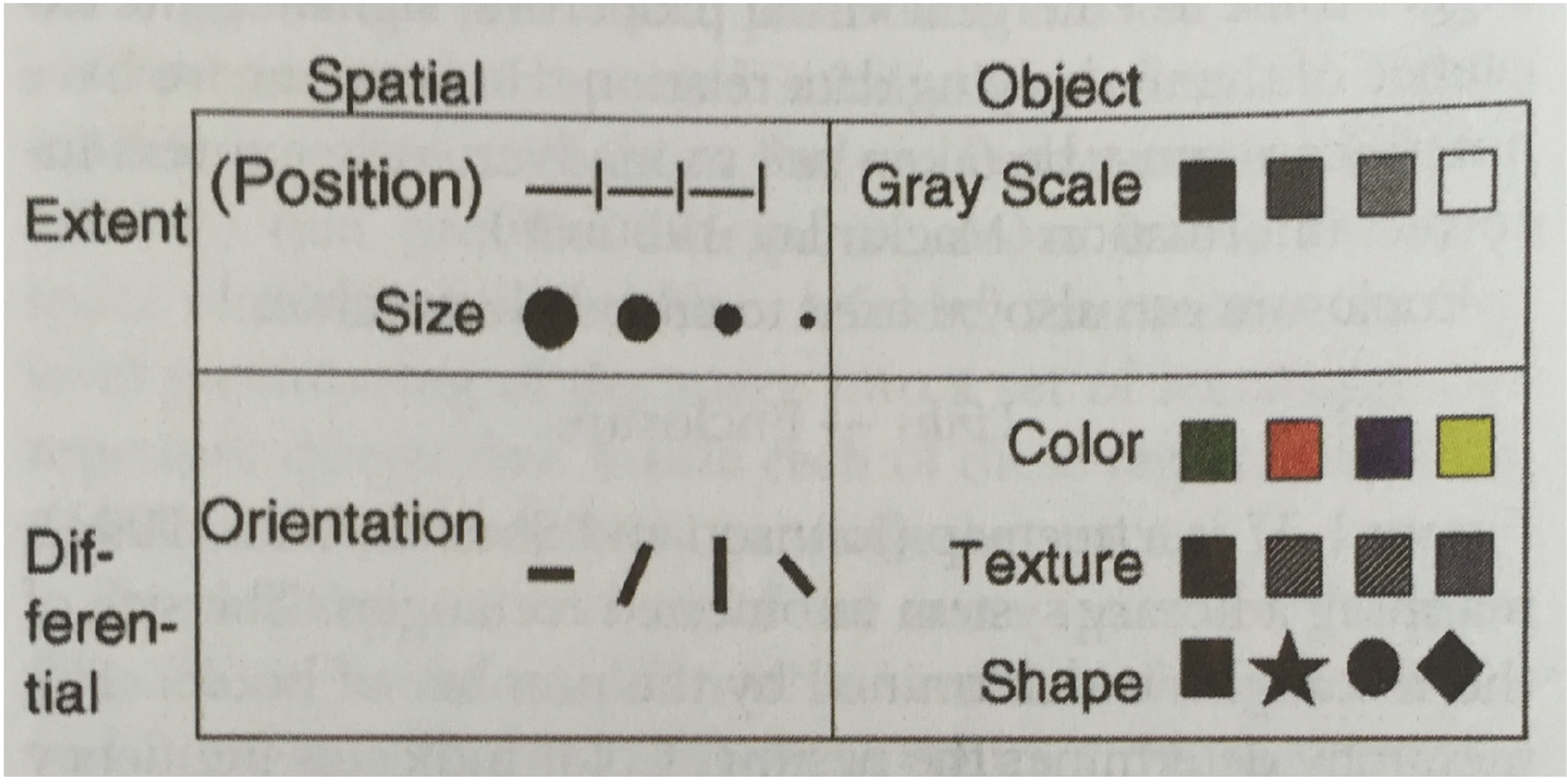


Marks

- Points (0D)
- Lines (1D)
- Areas (2D)
- Volumes (3D)

Marks' graphical properties

- Quantitative (Q), Ordinal (O), Nominal (N)
- Filled circle - good; open circle - bad



Effectiveness of graphical properties

- Quantitative (Q), Ordinal (O), Nominal (N)
- Filled circle - good; open circle - bad

	Spatial	Q	O	N	Object	Q	O	N
Extent	(Position)	●	●	●	Grayscale	◑	●	○
	Size	●	●	●				
Differential	Orientation	◑	◑	●	Color	◑	◑	●
					Texture	◑	◑	●
					Shape	○	○	●

Animation

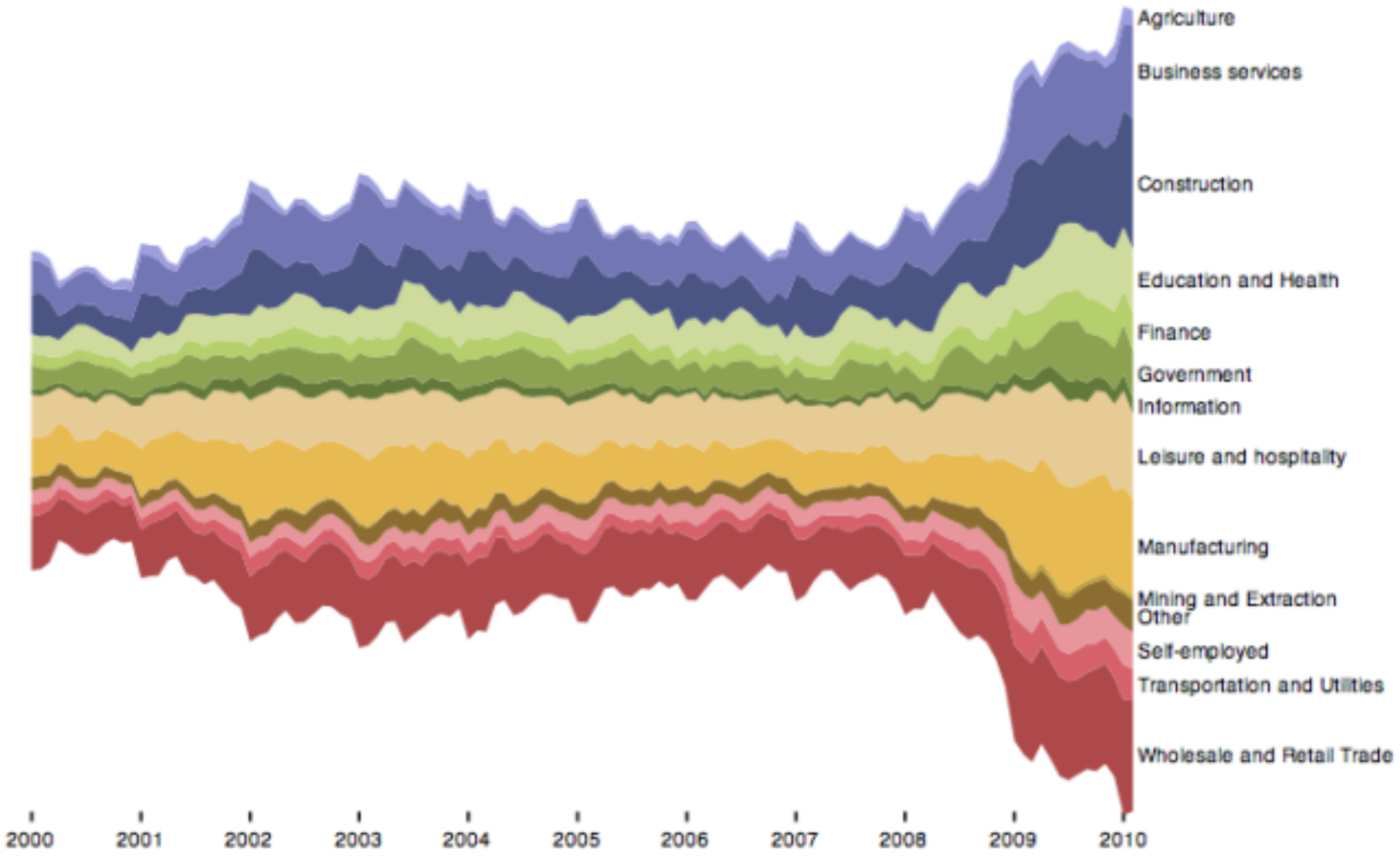
- Visualization can change over time
- Could be used to encode data as a function of time
  - But often not effective as makes direct comparisons hard
- Can be more effective to animate transition from before to after as user configures visualization

Examples of visualizations

Time-series data

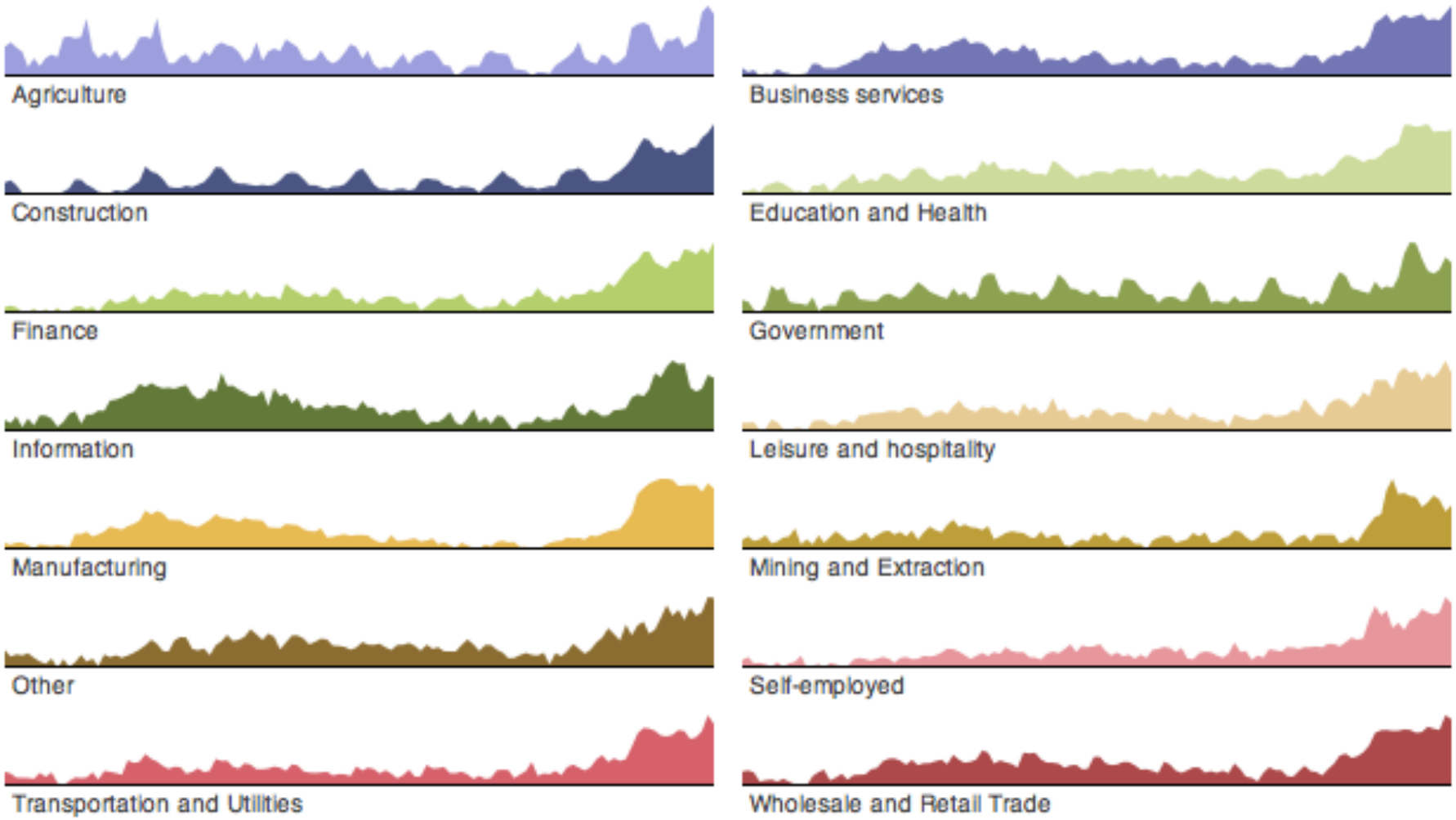
Stacked graph

- Supports visual summation of multiple components



Small multiples

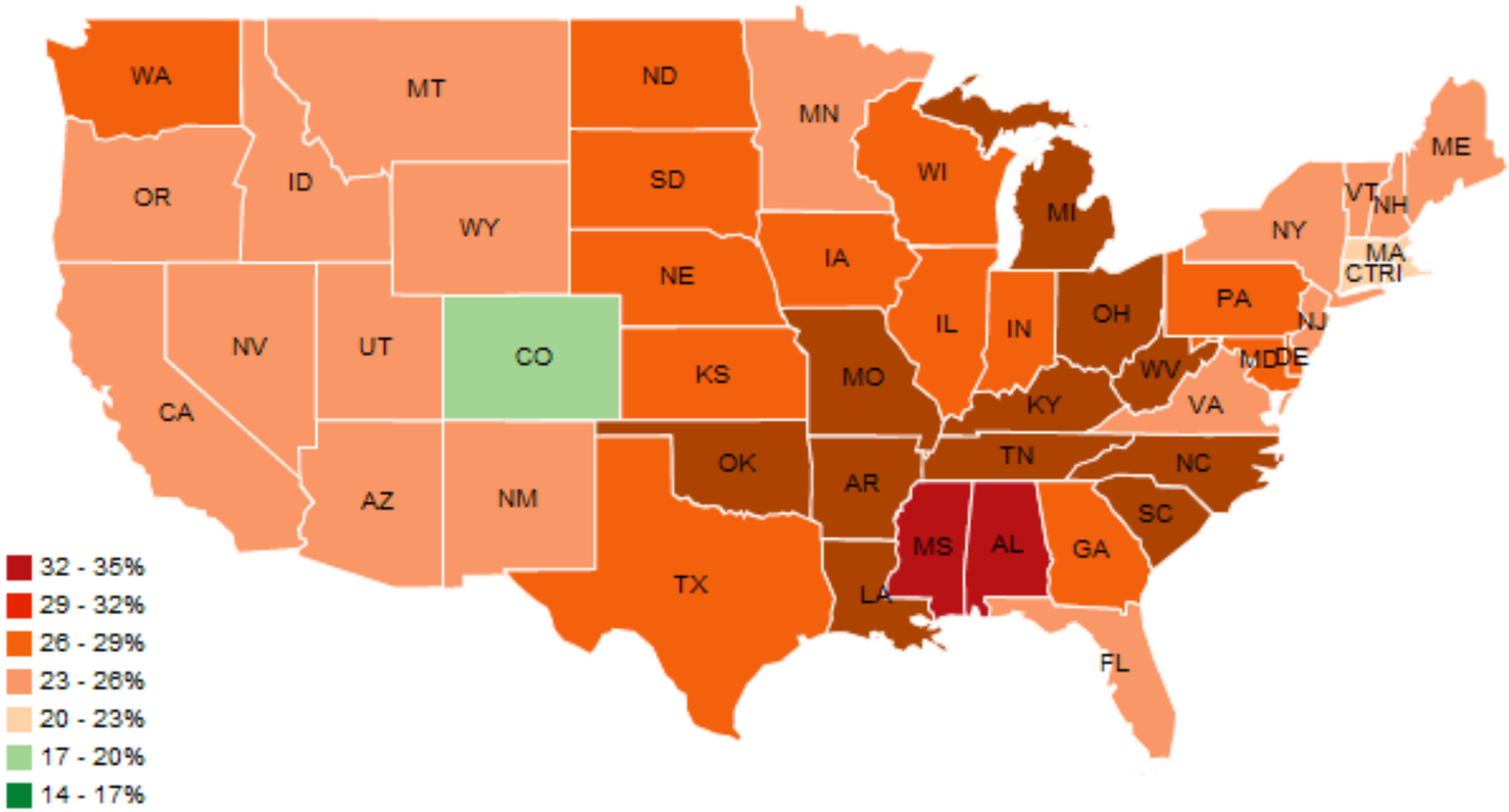
- supports separate comparison of data series
- may have better legibility than placing all in single plot



Maps

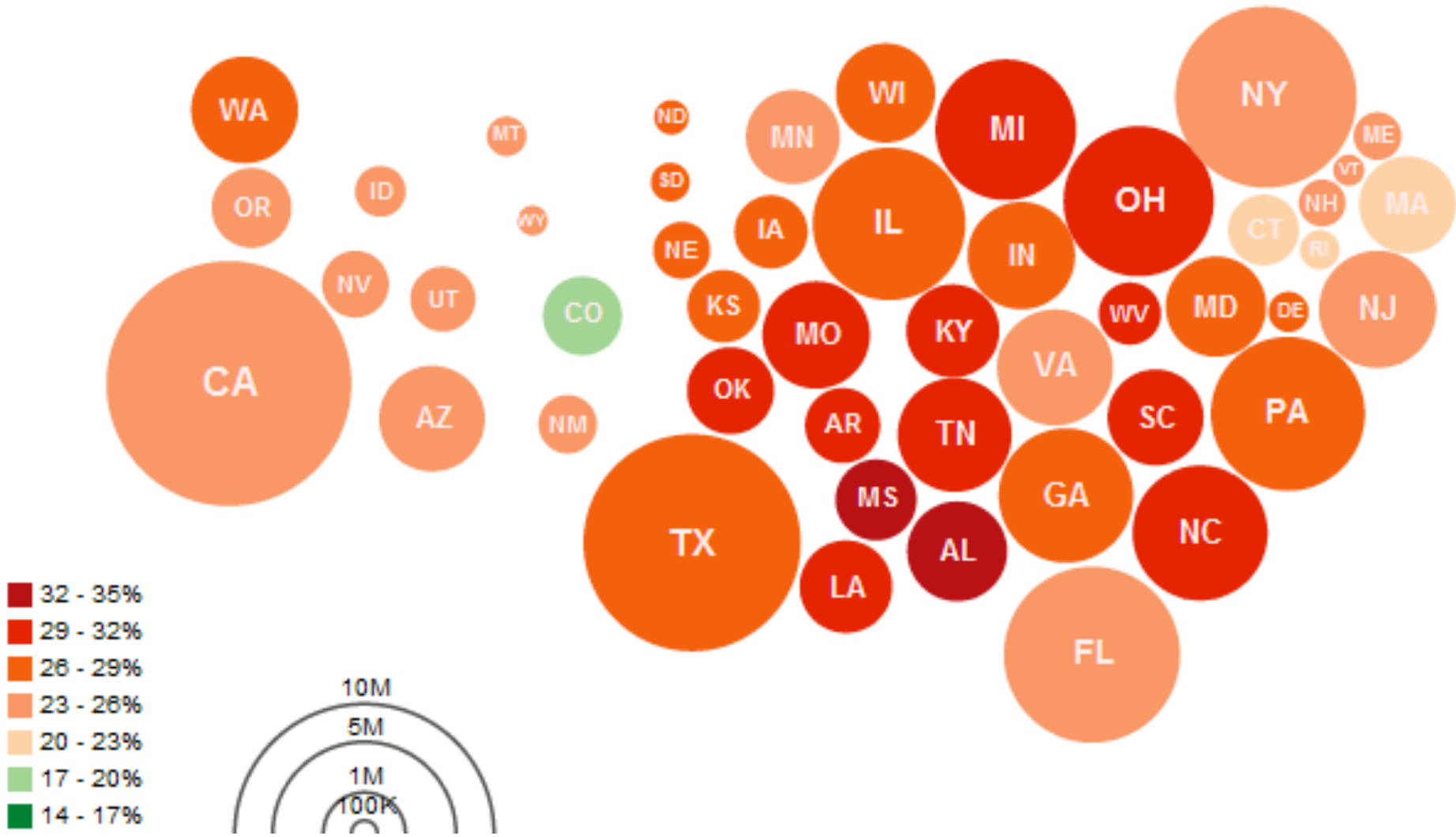
Choropleth map

- Groups data by area, maps to color



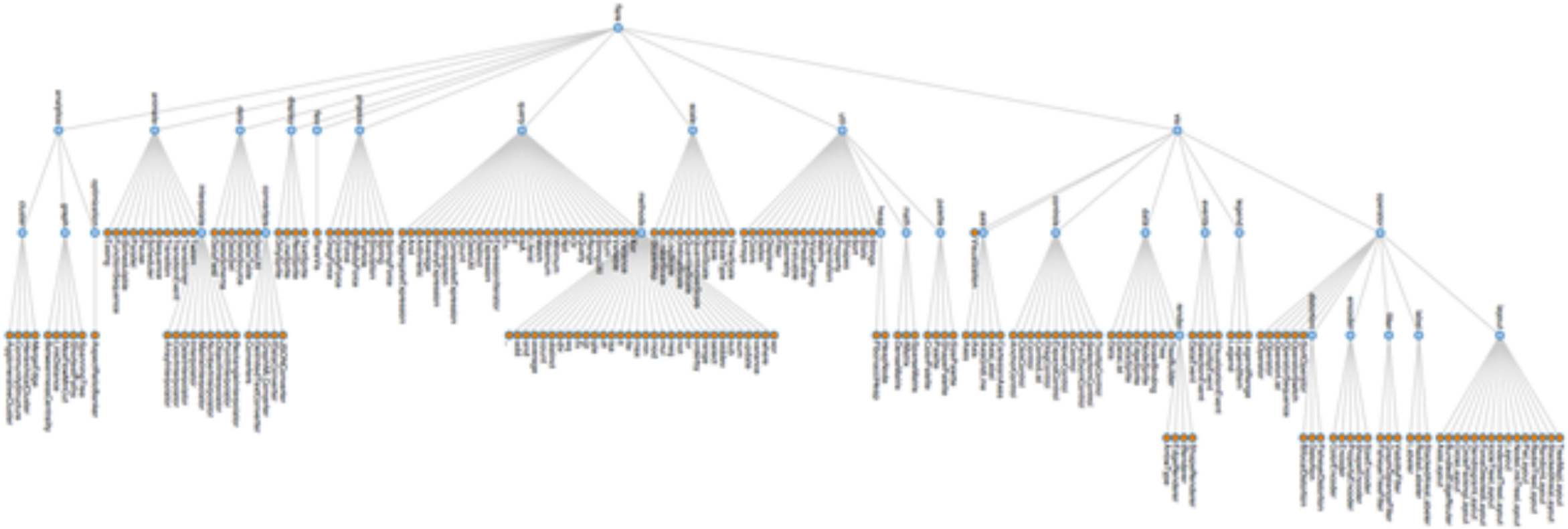
Cartograms

- Encodes two variables w/ size & color



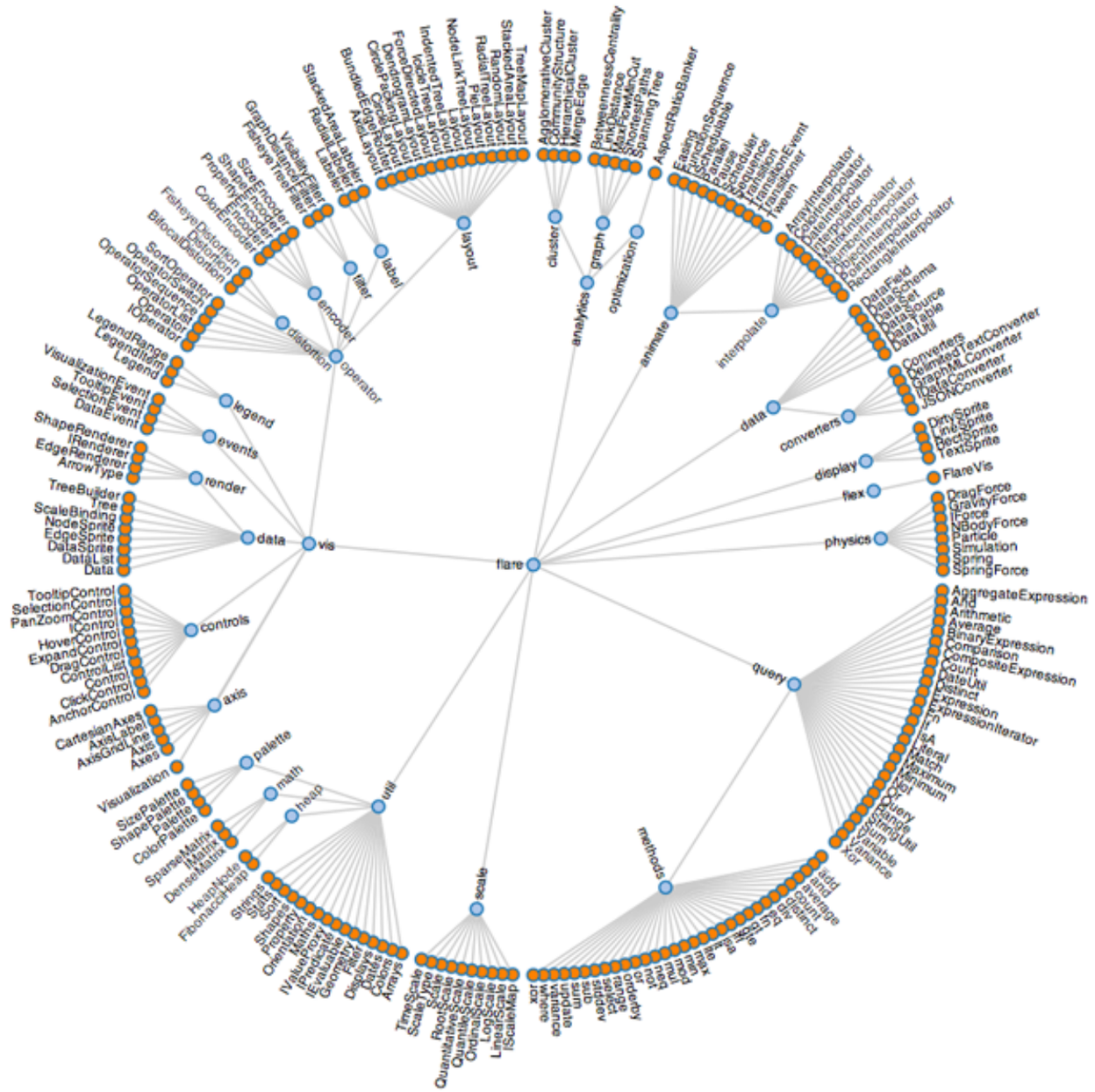
Hierarchies

Node link diagram

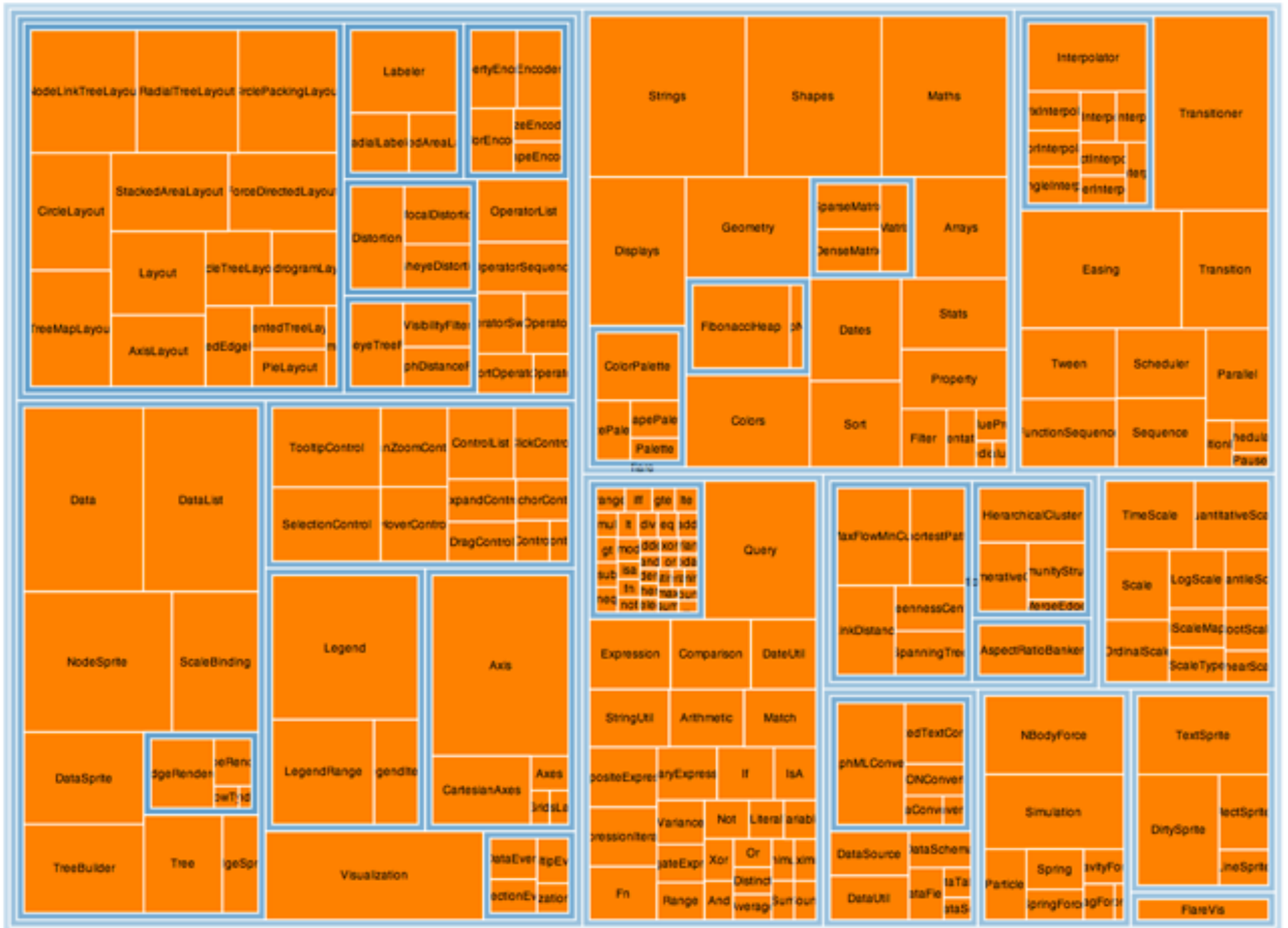


Dendrogram

- leaf nodes of hierarchy on edges of circle



Treemaps


Labeler
Strings
Shapes
Maths
Interpolator
Transitioner
Displays
Geometry
Arrays
Easing
Transition
Tween
Scheduler
Parallel
Data
DataList
Layout
CircleLayout
StackedAreaLayout
AxisLayout
PieLayout
Distortion
OperatorList
Colors
Sort
Dates
Stats
Property
Filter
Sequence
Pause
TooltipControl
ControlList
SelectionControl
DragControl
Query
Expression
Comparison
DateUtil
StringUtil
Arithmetic
Match
Legend
LegendRange
Axis
CartesianAxes
Axes
Visualization
Scale
LogScale
TimeScale
Simulation
NBodyForce
Spring
Particle
TextSprite
DirtySprite
NodeSprite
ScaleBinding
DataSprite
TreeBuilder
Tree
FlareVis

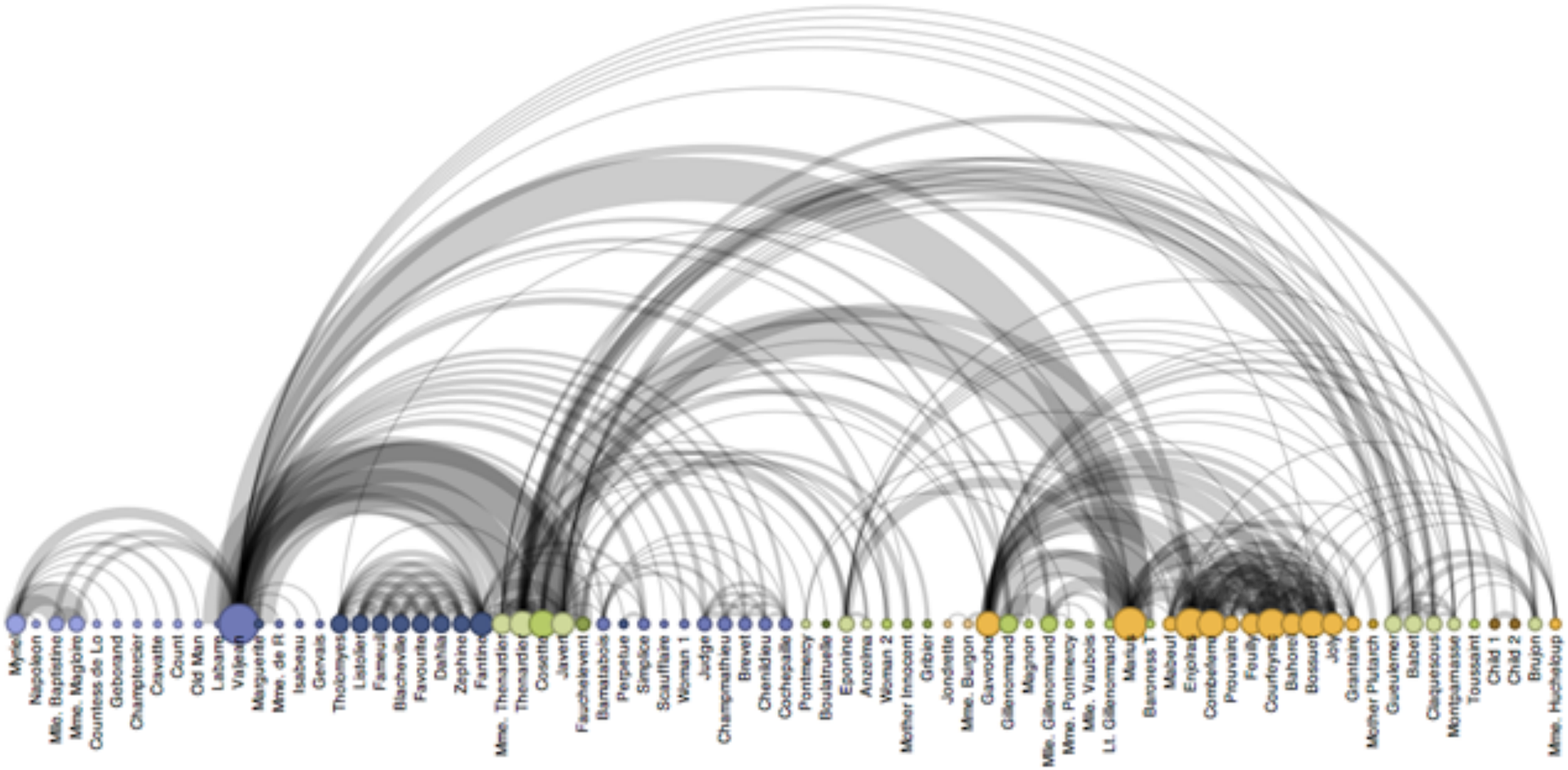
Networks

Force-directed layout

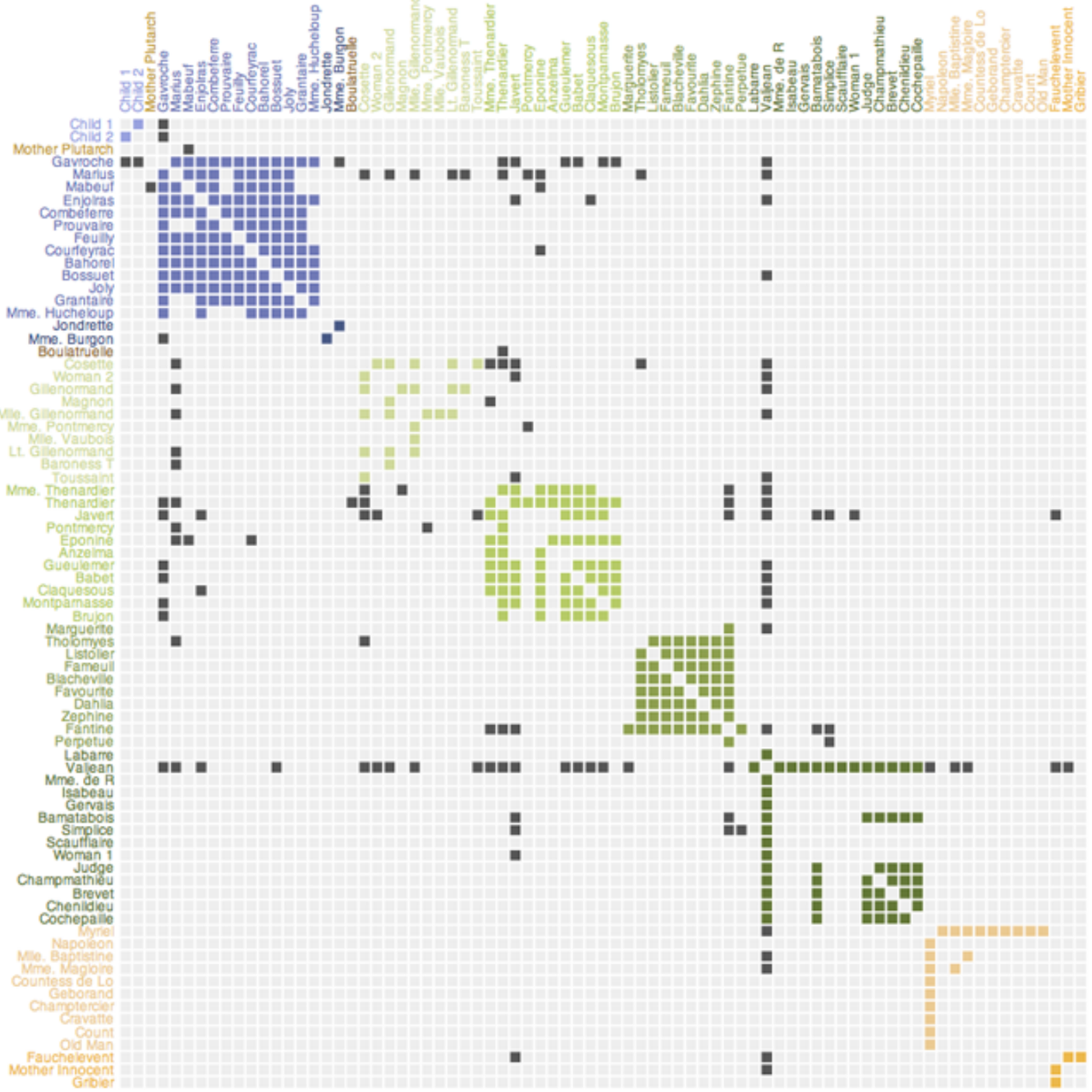
- edges function as springs, find least energy configuration

Arc diagram

- can support identifying cliques & bridges w/ right order



Adjacency matrix



Design considerations

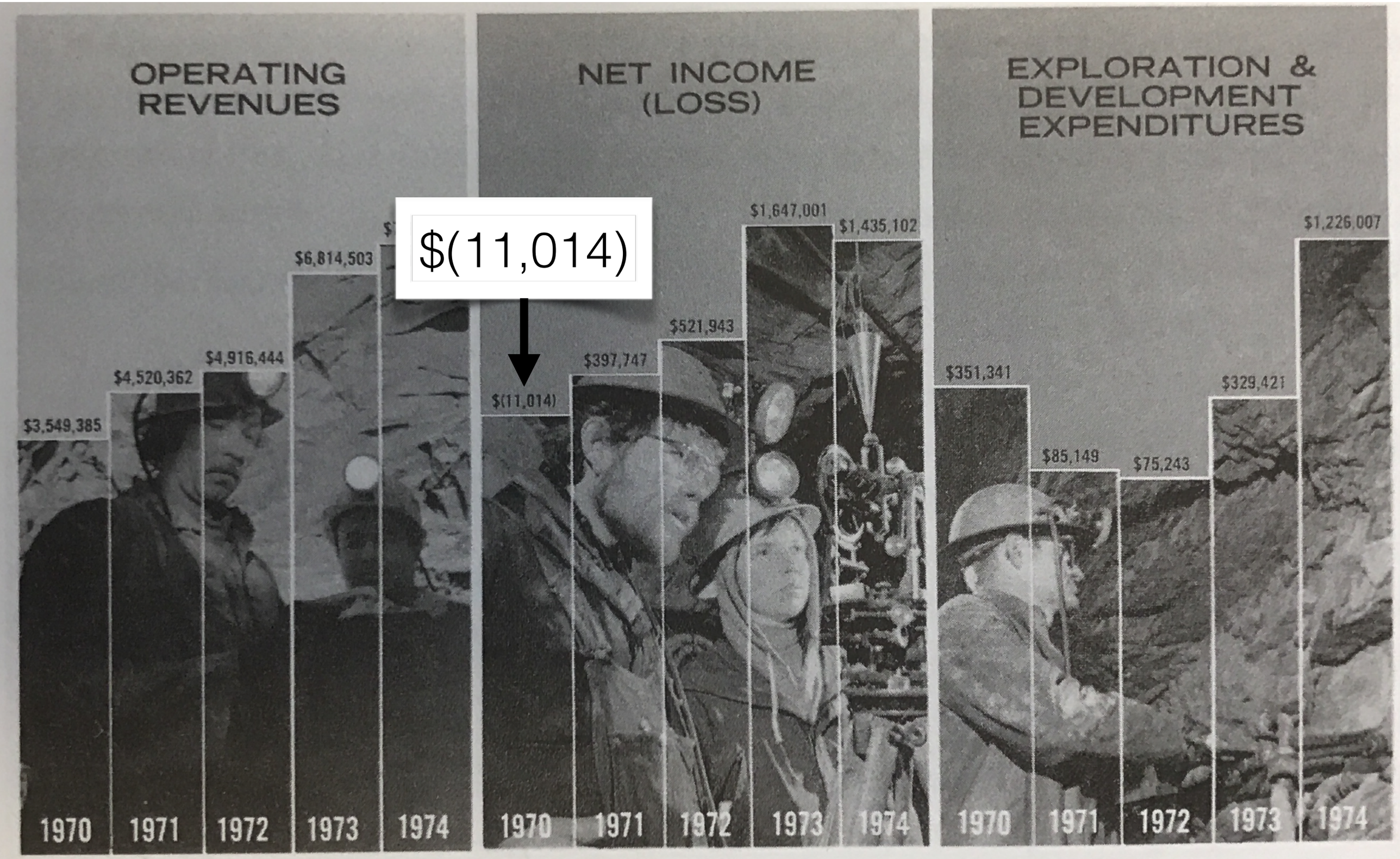
Tufte's principles of graphical excellence

- show the **data**
- induce the viewer to think about the substance rather than the methodology
- avoid distorting what the data have to say
- present **many** numbers in a small space
- make large data sets **coherent**
- encourage the eye to **compare** different pieces of data
- reveal data at several levels of detail, from overview to fine structure
- serve reasonable clear **purpose**: description, exploration, tabulation, decoration

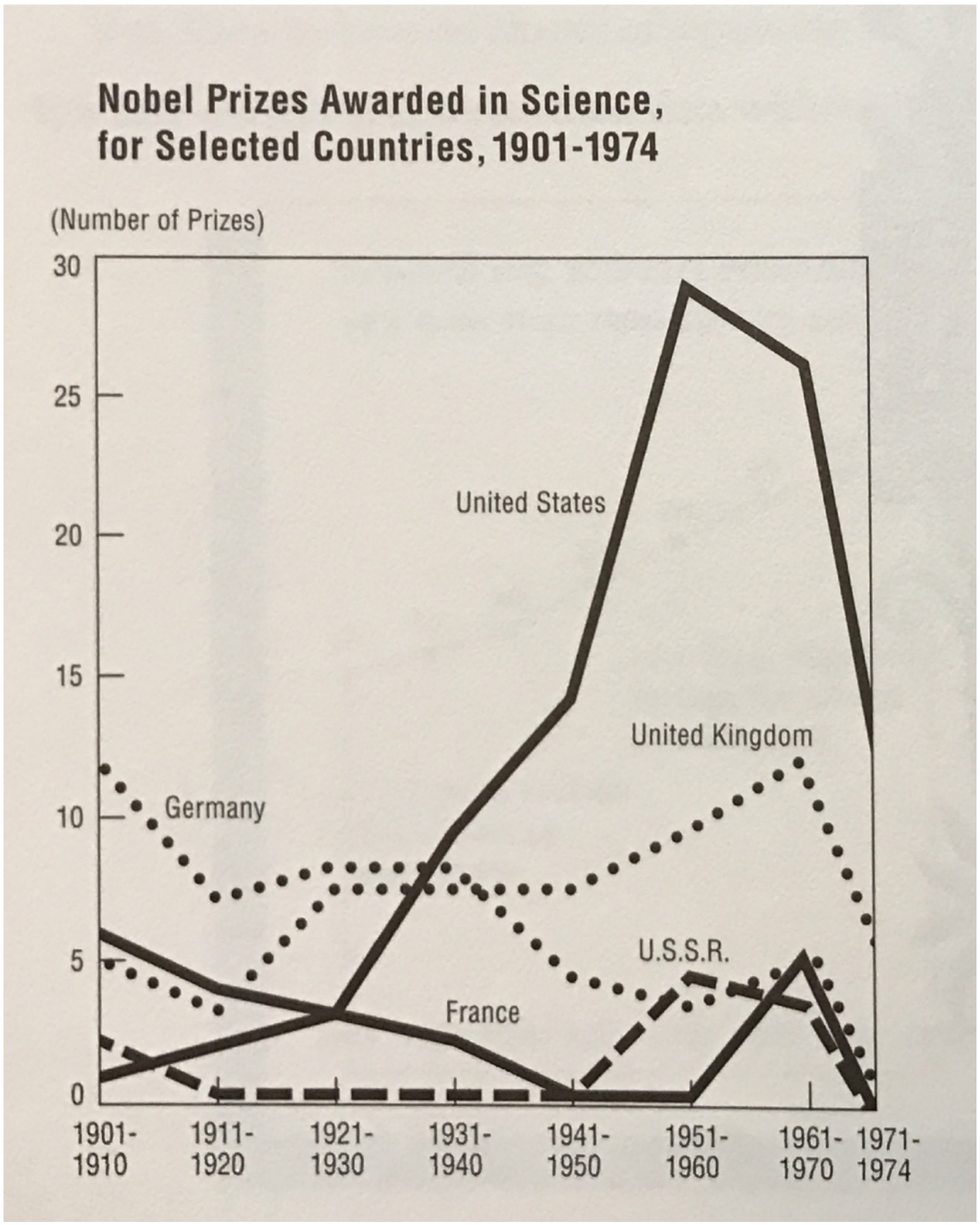
Distortions in visualizations

- Visualizations may distort the underlying data, making it harder for reader to understand truth
- Use of **design** variation to try to falsely communicate **data** variation

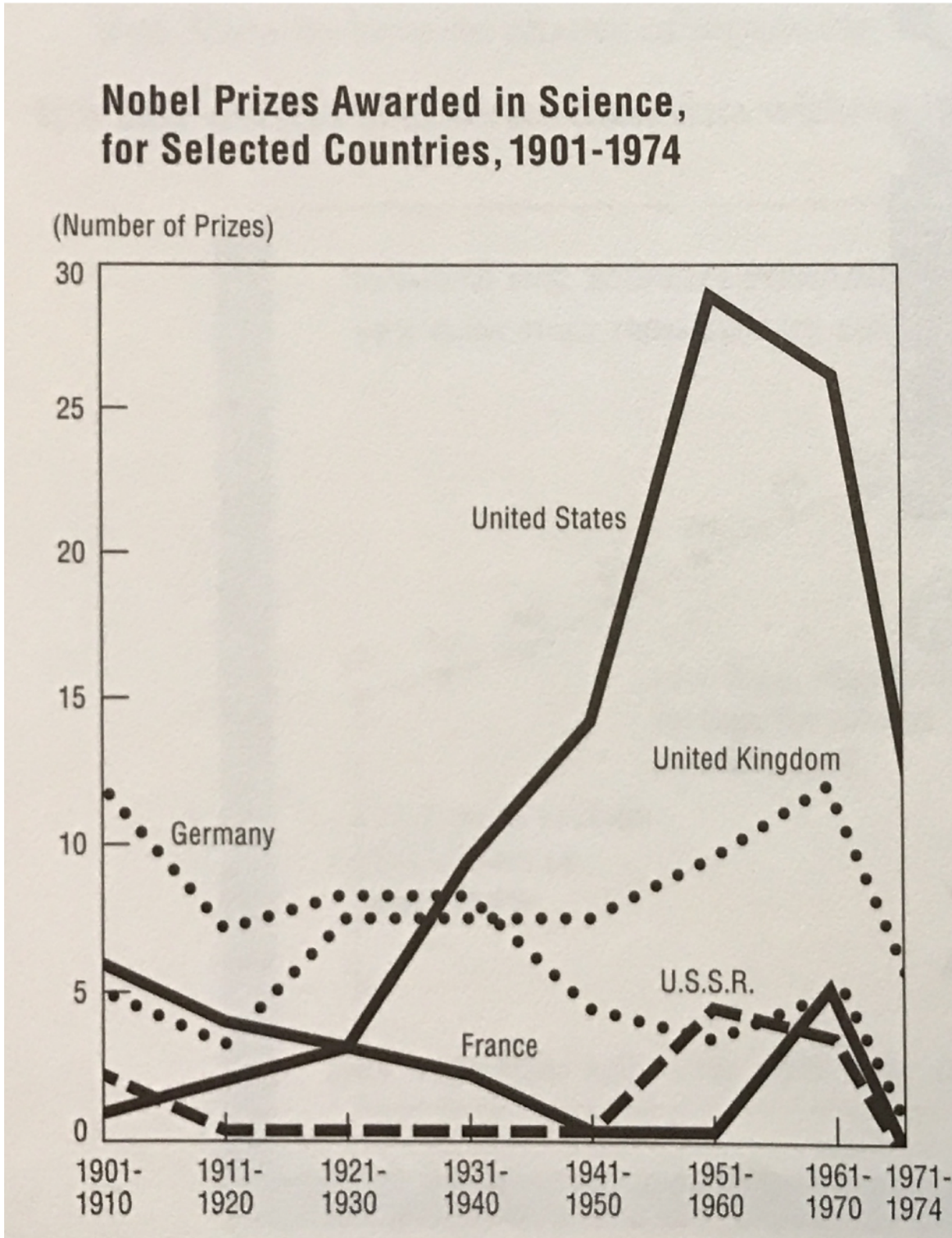
Example

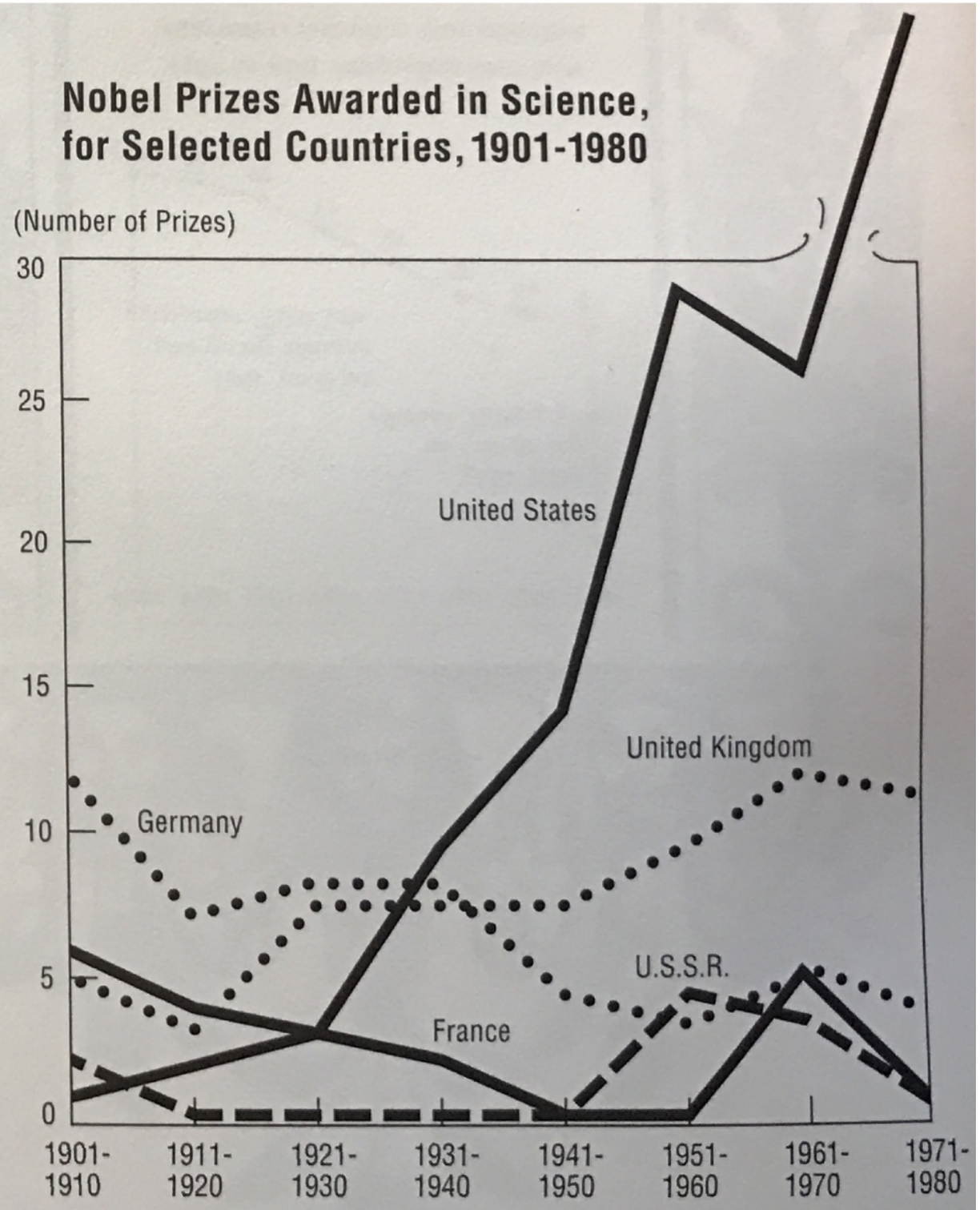

OPERATING REVENUES
$3,549,385
$4,520,362
$4,916,444
$6,814,503
1970
1971
1972
1973
1974
NET INCOME (LOSS)
$(11,014)
$(11,014)
$397,747
$521,943
$1,647,001
$1,435,102
1970
1971
1972
1973
1974
EXPLORATION & DEVELOPMENT EXPENDITURES
$351,341
$85,149
$75,243
$329,421
$1,226,007
1970
1971
1972
1973
1974

Example

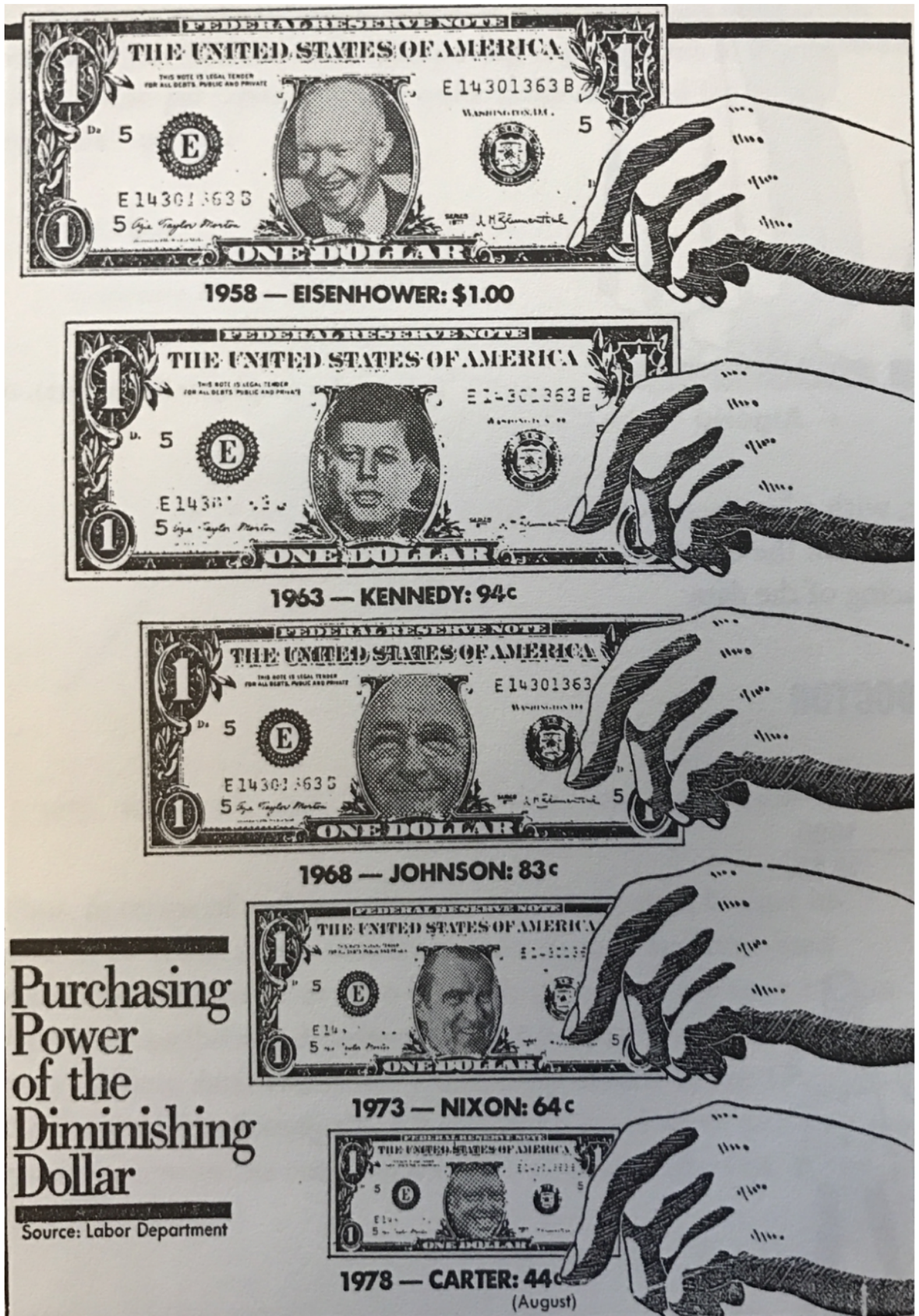

Nobel Prizes Awarded in Science,
for Selected Countries, 1901-1974
(Number of Prizes)
30
25
20
15
10
5
0
United States
United Kingdom
Germany
U.S.S.R.
France
1901-1910
1911-1920
1921-1930
1931-1940
1941-1950
1951-1960
1961-1970
1971-1974

Example (corrected)

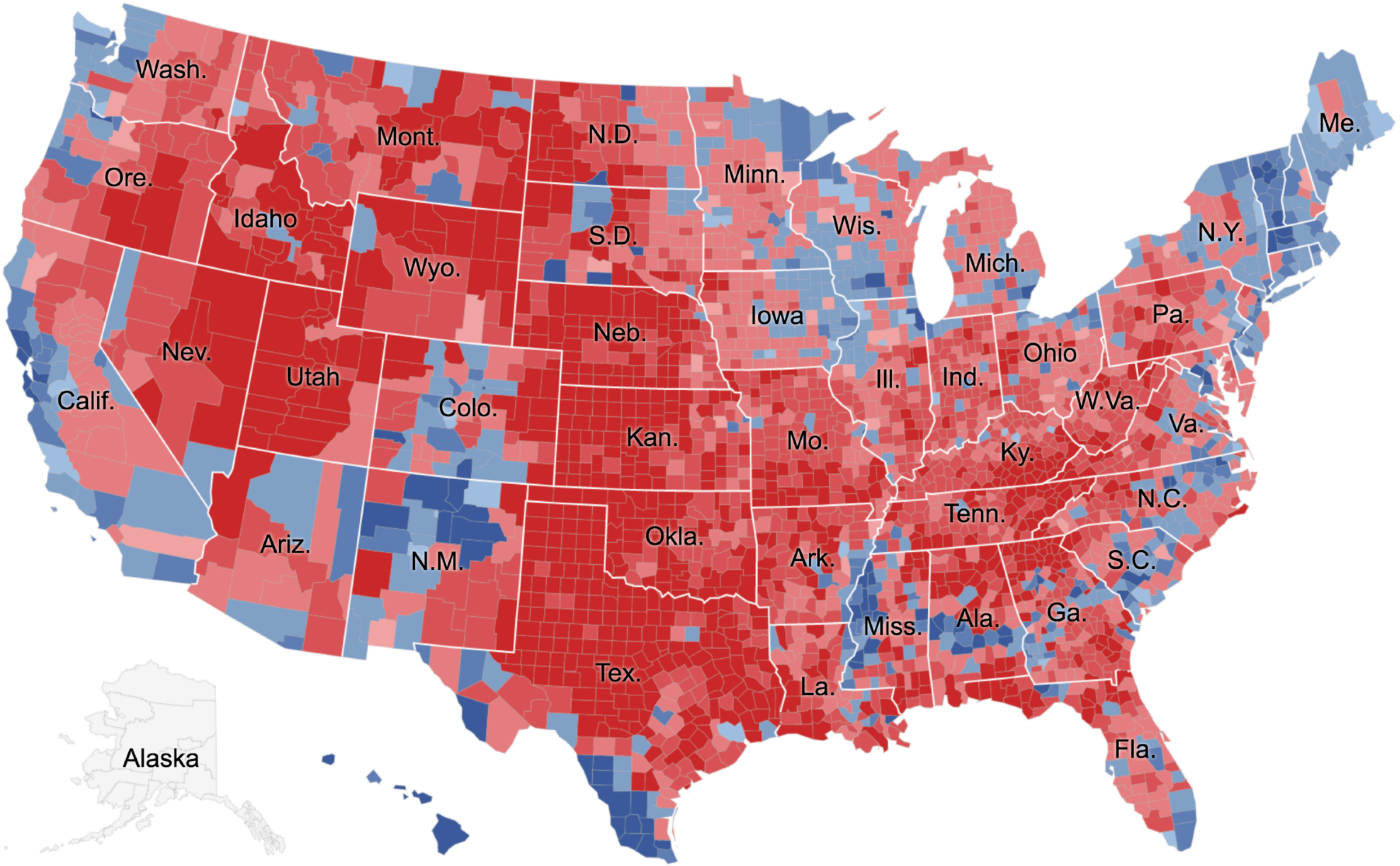




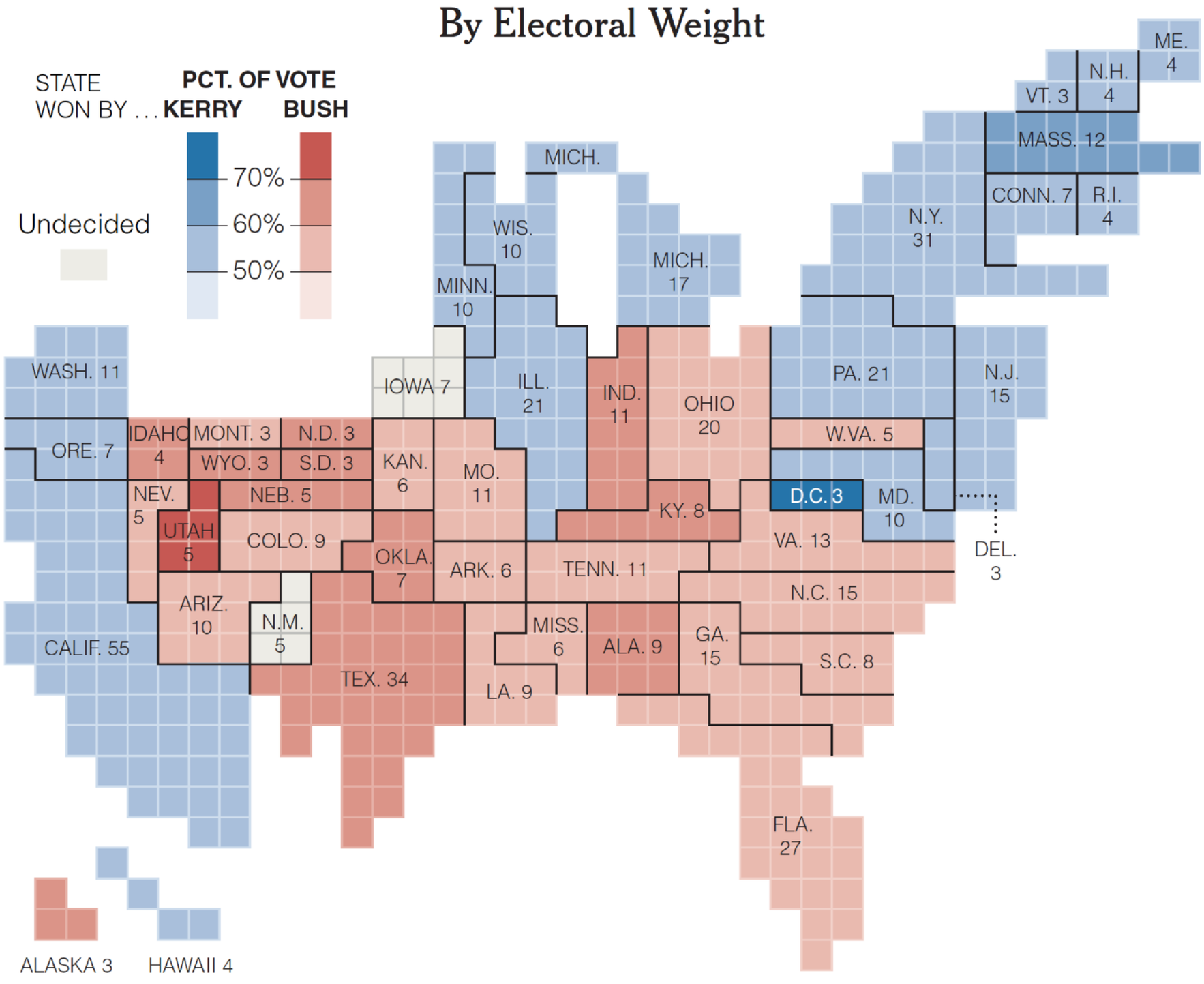
Example


1958 — EISENHOWER: $1.00
1963 — KENNEDY: 94c
1968 — JOHNSON: 83c
1973 — NIXON: 64c
1978 — CARTER: 44c
(August)
Purchasing Power of the Diminishing Dollar
Source: Labor Department

Traditional Electoral Map



Weighted Electoral Map



Data-ink

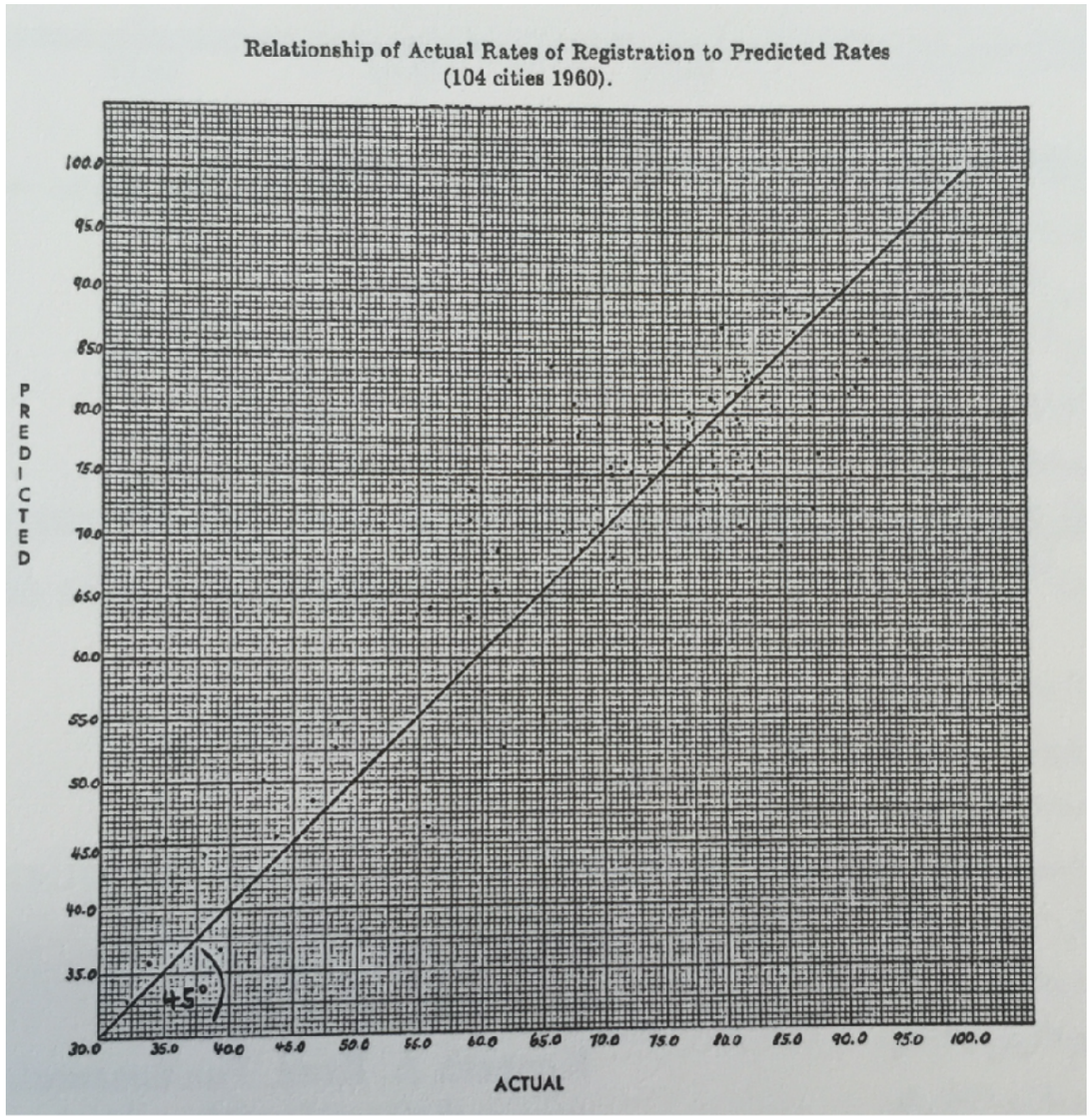
- Data-ink - non-redundant ink encoding data information

$$\text{Data-ink ratio} = \frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

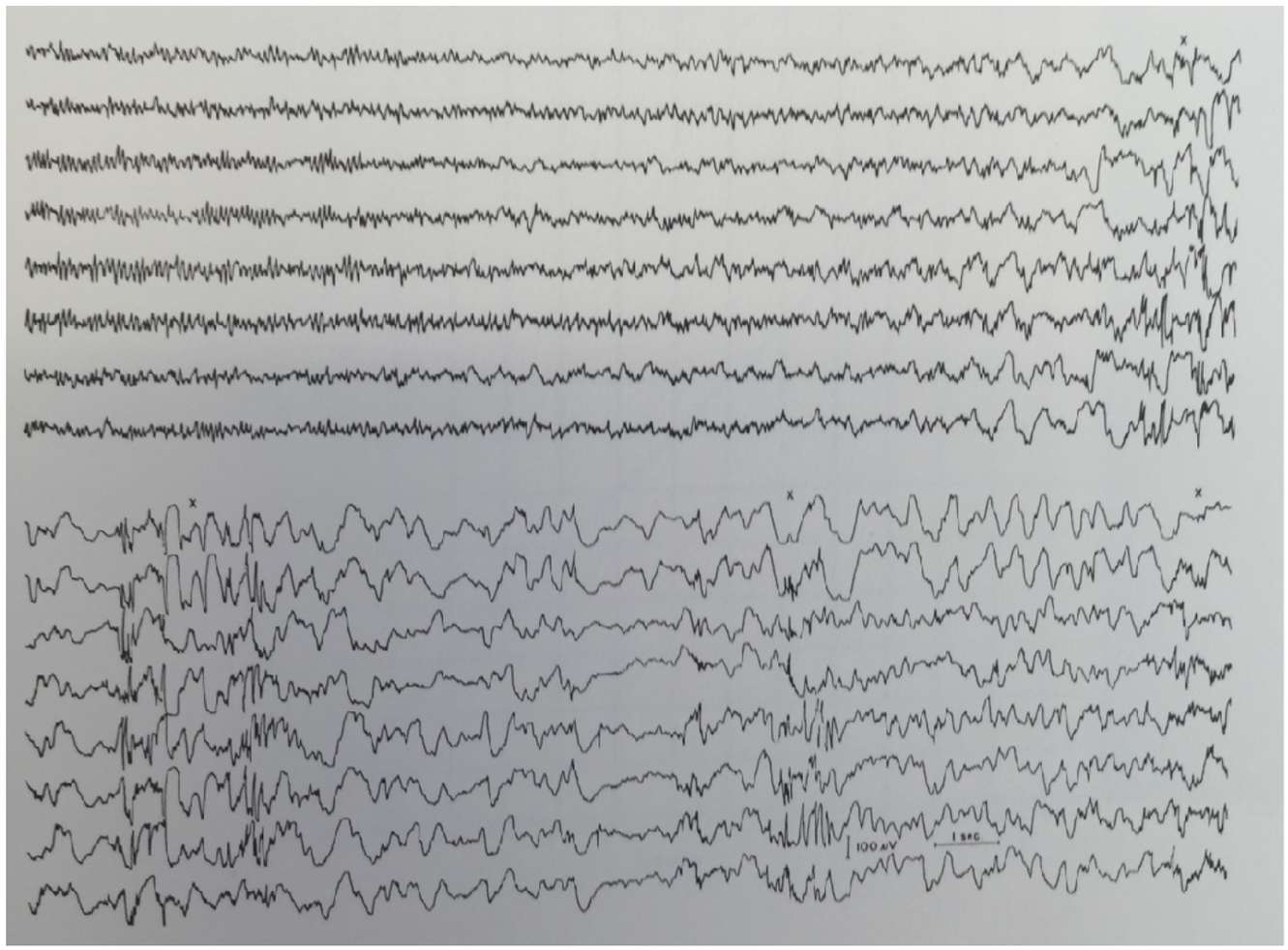
= proportion of a graphic's ink devoted to the non-redundant display of data-information

= 1.0 – proportion of a graphic that can be erased

Examples of data-ink ratio



~0

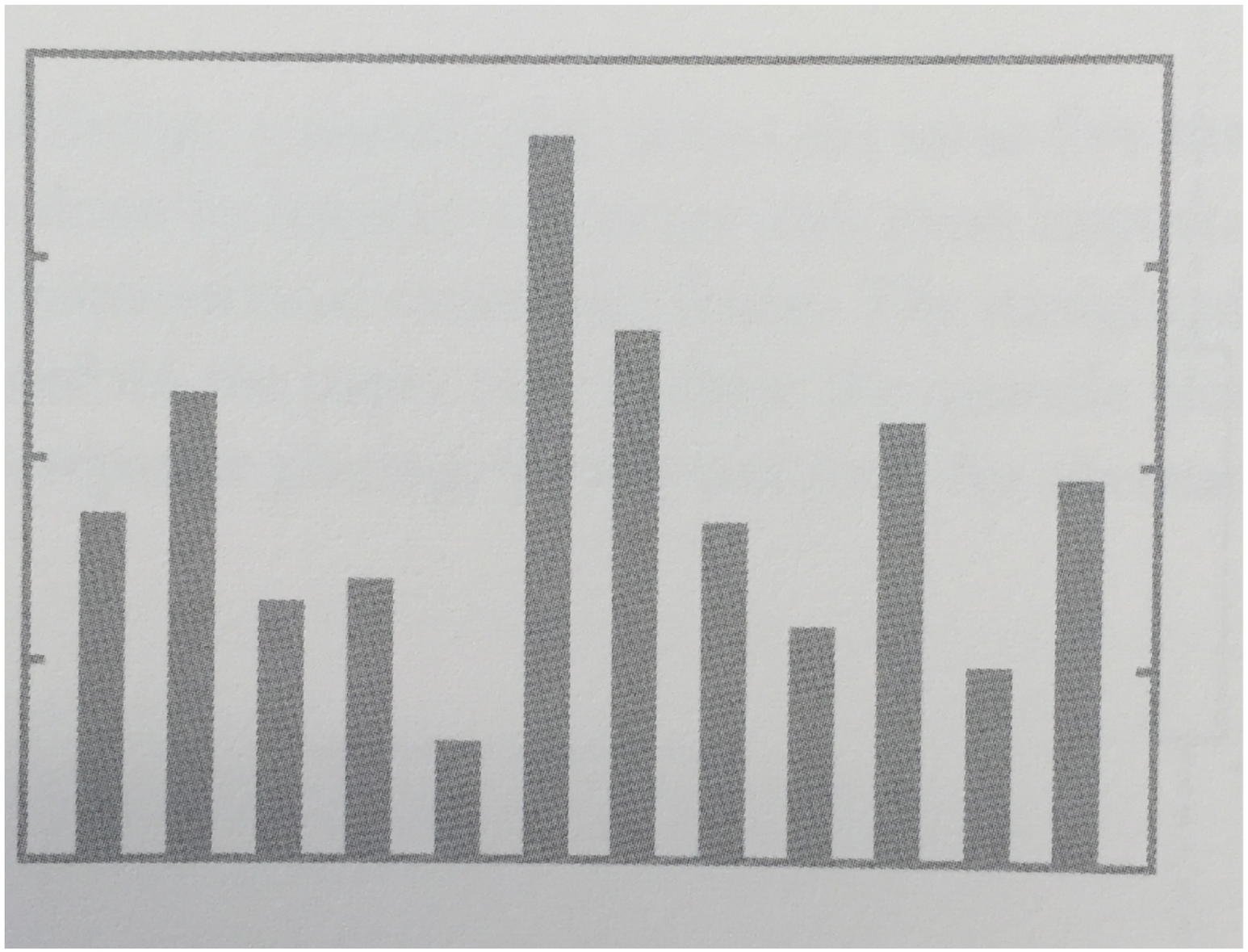


1.0

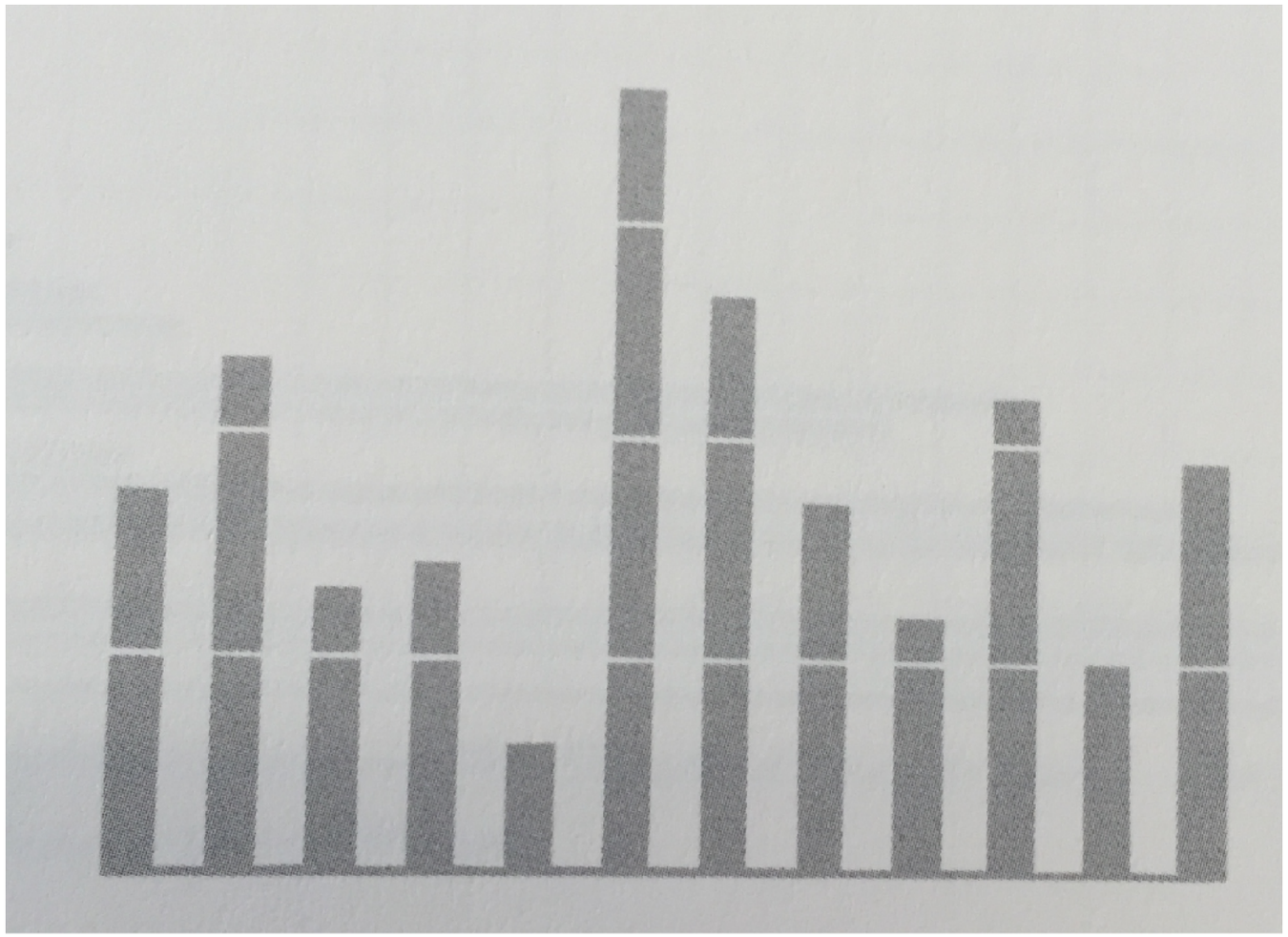
Design principles for data-ink

- (a.k.a. aesthetics & minimalism / elegance & simplicity)
- **Above all else show the data**
  - Erase non-data-ink, within reason
    - Often not valuable and distracting
    - Redundancy not usually useful

Example



Example (revised)



Interacting with visualizations

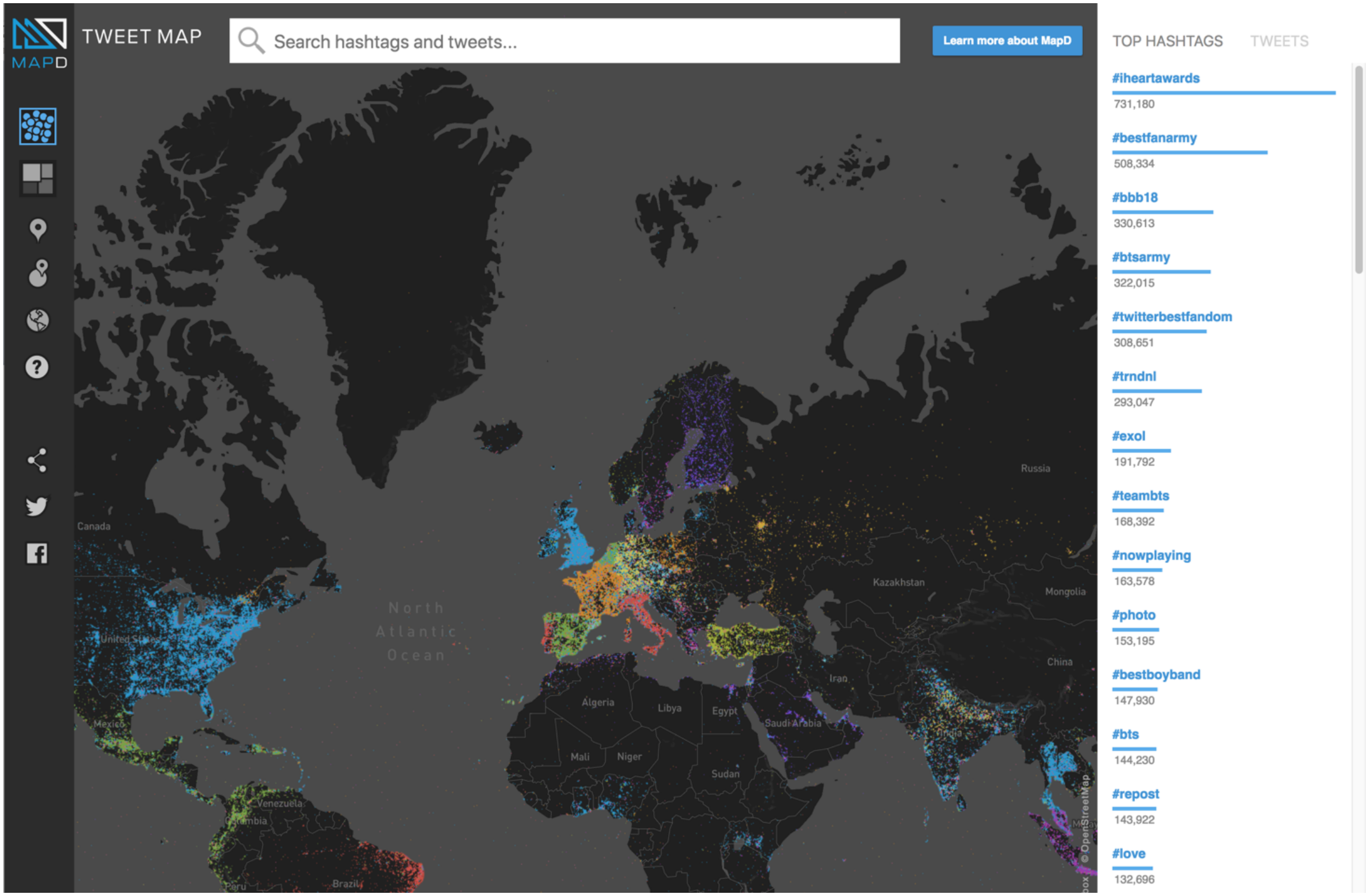
Interactive visualizations

- Users often use iterative process of making **sense** of the data
  - Answers lead to new questions
- Interactivity helps user constantly change display of information to answer new questions
- Should offer visualization that offers best view of data moment to **moment** as desired view **changes**

Information Visualization Tasks

- Overview: gain an overview of entire collection
- Zoom: zoom in on items of interest
- Filter: filter out uninteresting items
- Details on demand: select an item or group and get details
- Relate: view relationships between items
- History: support undo, replay, progressive refinement
- Extract: allow extraction of sub-collections through queries

Global tweet map



https://www.mapd.com/demos/tweetmap/

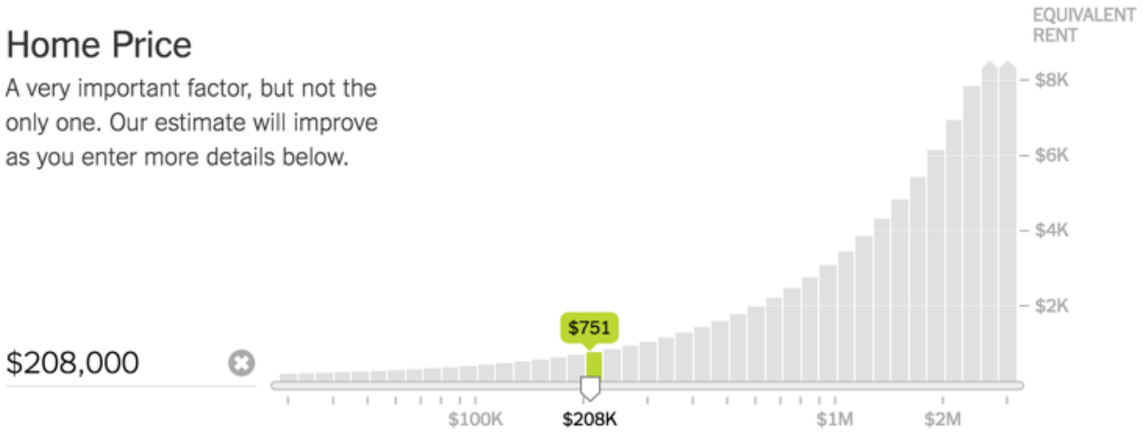
Is It Better to Rent or Buy?

By MIKE BOSTOCK, SHAN CARTER and ARCHIE TSE

The choice between buying a home and renting one is among the biggest financial decisions that many adults make. But the costs of buying are more varied and complicated than for renting, making it hard to tell which is a better deal. To help you answer this question, our calculator takes the most important costs associated with buying a house and computes the equivalent monthly rent. RELATED ARTICLE

Home Price

A very important factor, but not the only one. Our estimate will improve as you enter more details below.



If you can rent a similar home for less than ...

$751 PER MONTH

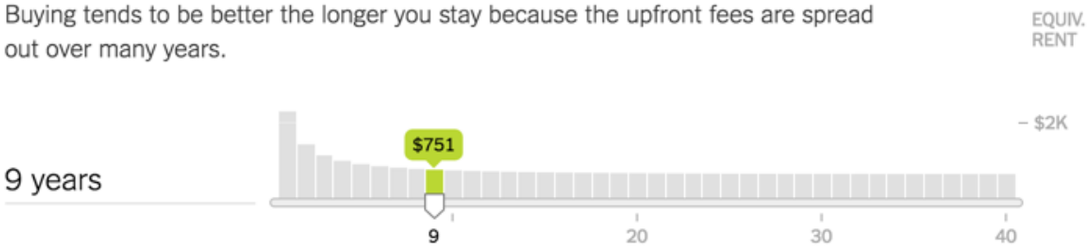
... then renting is better.

Costs after 9 years	Rent	Buy
Initial costs	$751	$49,920
Recurring costs	$90,937	$137,912
Opportunity costs	$13,092	$37,376
Net proceeds	-$751	-$121,180
Total	$104,029	$104,029

How to Read the Charts Charts that are relatively flat indicate factors that are not particularly important to the outcome. Conversely, the factors that have steep slopes have a large impact.

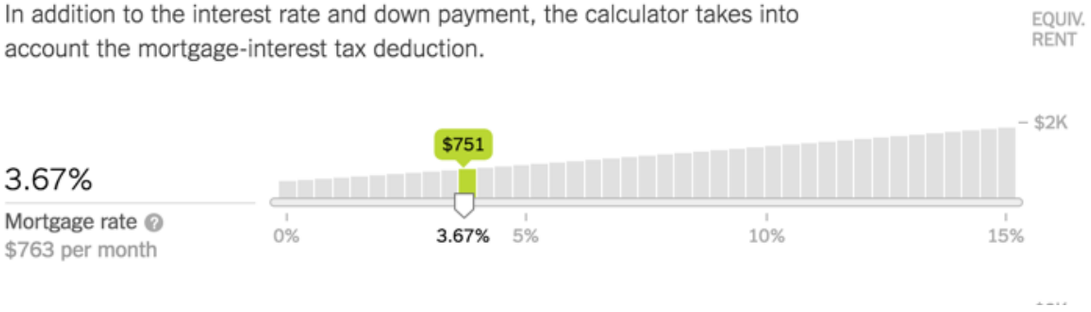
How Long Do You Plan to Stay?

Buying tends to be better the longer you stay because the upfront fees are spread out over many years.



What Are Your Mortgage Details?

In addition to the interest rate and down payment, the calculator takes into account the mortgage-interest tax deduction.



https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html?_r=0

In Class Activity

Design an information visualization

- In groups of 2 or 3
  - Select a set of data to visualize and two or more representative questions to answer using this data
  - Design an **interactive** information visualization
  - Create sketches showing the design of the information visualization
  - Should have multiple views of data, interactions to configure and move between views